Rec'd PCT/PTO 01 APR 2005

5//, 504

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷: H04B

(21) International Application Number: PCT/IB03/01187

(22) International Filing Date: 1 April 2003 (01.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02076649.9   25 April 2002 (25.04.2002)   EP

(71) Applicant (for all designated States except US): KONIN-KLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and
(75) Inventor/Applicant (for US only): DE OLIVEIRA KAS-TRUP PEREIRA, Bernardo [BR/NL]; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: DUIJVESTIJN, Adrianus, J.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/092171 A2

(54) Title: PROCESSING METHOD AND APPARATUS FOR IMPLEMENTING SYSTOLIC ARRAYS

(57) Abstract: The present invention relates to a processing method and apparatus for implementing a systolic-array-like structure. Input data are stored in a depth-configurable register means (DCF) in a predetermined sequence, and are supplied to a processing means (FU) for processing said input data based on control signals generated from instruction data, 5 wherein the depth of the register means (DCF) is controlled in accordance with the instruction data. Thereby, systolic arrays can be mapped onto a programmable processor, e.g. a VLIW processor, without the need for explicitly issuing operations to implement the register moves that constitute the delay lines of the array.

Processing method and apparatus for implementing systolic arrays

The present invention relates to a processing method and apparatus, especially a scaleable VLIW (Very Large Instruction Word) processor or a coarse-grained reconfigurable processor for implementing systolic-array-like structures.

Programmable or configurable processors are pre-fabricated devices that can be customised after fabrication to perform a specific function based on instructions or configurations, respectively, issued to it. These instructions or configurations, when executed in the processor, control the processor resources (e.g. arithmetic logic unit (ALU), register file, interconnection, memory, etc.) to perform certain operations in time (i.e. sequentially) or space (i.e. in parallel). Typically, configurable processors will perform more operations in space than programmable processors, while programmable processors will perform more operations in time than configurable processors.

An algorithm-to-silicon design methodology for digital signal processors (DSP) has been developed, which allows for an enormous increase in design productivity of a DSP designer and a more optimised design of the resulting chip. The methodology initially involves capturing an algorithm in an implementation-independent way. Then, with the help of a set of evaluators and analysers, the algorithm can be tuned and optimised for a fixed-point implementation. Once a satisfactory behaviour is reached, a set of interactive synthesis engines can be applied to map the fixed-point specification to a target VLIW-like architecture. This mapping process is very flexible and fast, which makes it possible to try out many alternatives in a very short time. In general, a very large instance of such a VLIW-like processor architecture can be seen as a coarse-grained reconfigurable processor, in which each control word in its micro-code memory is a configuration. This interpretation is possible due to the size of the corresponding VLIW instruction, which allows for many parallel operations to be performed, therefore largely computing in space.

VLIW processors are used to exploit the available Instruction Level Parallelism (ILP) in an application. To exploit the ILP, data-independent operations are scheduled concurrently in a VLIW instruction.

Fig. 1 shows a schematic diagram indicating a processing application and a corresponding programmable processor structure of an application, where a data-flow graph

representing a loop body is shown on the left side. In Fig. 1, circles 20 represent operations, and arrows represent data dependencies between operations. Dashed arrows represent input or output values respectively consumed or produced in a loop iteration. On the right-hand side, a 4-issue slot VLIW processor 10 is depicted, comprising four ALUs A1 to A4 and four issue slots I1 to I4 for controlling the operation of the ALUs A1 to A4. In the present case, the VLIW processor 10 can compute one iteration of the indicated loop processing application in five cycles, executing a sequence of two, four, two, one, and one operation(s) in each cycle, respectively. The number of operations per cycle depends on the number of operations which can be processed concurrently or in parallel, i.e. shown within one horizontal line of the processing application. The partial area 30 of the processing application illustrates the situation in the second cycle, in which four operations are executed in parallel, in one cycle of the VLIW processor 10.

Note that ILP is exploited within the loop body for a single iteration of the loop. Techniques of software pipelining can be used to exploit ILP across loop iterations, but those are typically difficult to implement and are mostly effective only for very simple and small loops, e.g. single basic blocks.

Custom hardware, however, can overlap the execution of every iteration of a loop, keeping most computing resources busy at all cycles. This kind of implementation exploits data locality and pipelining to the extreme. It is known as systolic arrays. Fig. 2 shows a schematic diagram indicating a systolic array implementation of the final two taps of a digital filter application, e.g. an FIR (Finite Impulse response) filter which can generate an output sample at every cycle. The grey blocks are clocked registers R. All function units FU are also busy at every cycle. Input data i is processed locally as it goes down the "pipe" to the right to generate output data o, as in a "pulsating" assembly line. The line acc contains partial accumulations. Registers c contain coefficients to the multipliers. Therefore, this architecture is called "systolic" array. Systolic arrays allow for very high exploitation of parallelism, obtaining high throughput.

In Zapata et al, "A VLSI constant geometry architecture for the fast Hartley and Fourier transforms", IEEE Transactions on Parallel and Distributed Systems, Vol. 3, No. 1, pp. 58-70, January 1992, an organization of a processor memory is based on first-in-first-out (FIFO) queues to facilitate a systolic data flow and to permit implementation in a direct way of complex data movements and address sequences of the transforms. This is accomplished by means of simple multiplexing operations, using hardware control.

Hence, it is, in principle, possible to map systolic arrays onto a VLIW processor. Then, each function unit FU in the systolic array will correspond to an equivalent unit (e.g. ALU, multiplier, MAC, etc.) in the VLIW processor and will be allocated one issue slot. For the systolic array of Fig. 2, four issue slots would be required in the VLIW processor for the four function units FU. In addition, one register move unit would be required in the VLIW processor for each register move corresponding to a delay line in the systolic array, with its corresponding issue slot. In the systolic array of Fig. 2, seven register moves that correspond to delay lines are provided. Therefore, seven register move units would be required in the VLIW processor, with their additional seven issue slots. This way, there would be more issue slots and, therefore, control signals and associated circuitry, corresponding to register moves than to actual operations. Also, the need for the move units to access the same registers that need to be accessed by other function units, introduces architectural complications in the VLIW design. All this renders the VLIW implementation of systolic arrays impractical. In this respect, it is noted that, in the original systolic array, register moves are encoded in space, by means of FIFO lines of registers that can implement the delay lines without any explicit control.

It is an object of the present invention to enable implementation of systolic array structures by a programmable processor.

This object is achieved by a processing apparatus as claimed in claim 1 and by a processing method as claimed in claim 8.

Accordingly, a programmable processor template for implementing systolic arrays can be achieved by providing a depth-configurable register means at the input of the processing units. Due to the possible implementation of systolic array structures by programmable processors, e.g. VLIW processors, hardware-like performance, mainly throughput, can be provided for media intensive applications, like video streaming, while preserving the flexibility and programmability of a well-known processor paradigm. It could even be possible to get a compiler to automatically generate "systolic array-like" instruction schedules, without need for explicit hardware design. Compilation technology could be extended in this direction.

Thus, a cost-effective VLIW template can be provided for the mapping of systolic structures. This template considerably reduces the overhead created by the current need to explicitly control all register move operations corresponding to delay lines

Preferably, the register means may comprise distributed register files provided at each input terminal of a plurality of functional units of the processing means. In particular,

the distributed register files may comprise depth-configurable FIFO register files addressable for individual registers. The number of physical registers available is fixed by the hardware. Then, the register control means may be arranged to determine the last logical register of the FIFO register files based on control signals derived from the instruction data.

Furthermore, at least one issue slot may be provided for storing the instruction data. The register control means may be arranged to use a part of the bit pattern of the instruction data stored in the at least one issue slot for controlling the depth of the register means.

Other advantageous further developments are defined in the dependent claims.

In the following, the present invention will be described on the basis of a preferred embodiment with reference to the accompanying drawings in which:

Fig. 1 shows a schematic diagram of a processing application and a corresponding programmable processor structure;

Fig. 2 shows a schematic diagram of a systolic array architecture;

Fig. 3 shows a principle architecture for implementing the systolic array architecture of Fig. 2 in a programmable processor according to the present invention; and

Fig. 4 shows a programmable processor architecture according to the preferred embodiment for implementing systolic arrays.

The preferred embodiment will now be described on the basis of a VLIW processor architecture.

In Fig. 3, the systolic array of Fig. 2 is restructured to enable its implementation in a VLIW architecture. Issue slots I1 to I4 are made explicit, and first-in-first-out (FIFO) delay lines comprising registers R are preserved at the input terminals of functional units FU, e.g. ALUs. Dotted boxes represent physical registers available in the hardware but not used in the shown systolic configuration. Drawn this way, the scheme suggests a VLIW template that can efficiently map systolic structures. The intuitive concept illustrated in Fig. 3 can be generalised by providing distributed register files at each input of the functional units FU.

Fig. 4 illustrates a programmable processor architecture according to the preferred embodiment as a VLIW template which can efficiently map systolic structures. In

particular, distributed register files DCF are provided, one for each input of each function
unit FU. Additionally, an interconnect network consisting of several point-to-point lines is
provided and connected to the respective inputs of the functional units by input or output
multiplexers 50. Thereby, the point-to-point lines can be written to by a single predetermined
function unit FU. Although Fig. 4 suggests full connectivity, the interconnection bus does not
need to be fully connected. Furthermore, each input of a functional unit FU can be connected
to a standard register file RF, addressable for individual registers. Note that in Fig. 4, for
simplicity, only the right one of the inputs of each functional unit FU is shown connected to a
respective standard register file RF. Register files with multiple read and/or write ports are
also possible.

        Due to the fact that the template does not include any centralised structure, i.e.
all resources are distributed, it is scaleable, allowing for very high number of issue slots
potentially needed by large systolic arrays, e.g. a 16-tap FIR filter or a large matrix
multiplier.

        According to the preferred embodiment, a depth-configurable register file
DCF is arranged at each input of each function unit FU. The depth-configurable register files
DCF may be implemented by FIFO memories whose last logical register can be determined
by control signals. However, any other addressable or controllable memory or register
structure capable of determining a last logical storage position in a delay line based on
control or address signals can be used for implementing the depth-configurable register files
DCF.

        For a depth-configurable FIFO of N physical registers, the output of the FIFO
can be programmed to be at register N, N-1, N-2, ... 1. By controlling the depth of the FIFO,
we can control the number of delay lines it emulates. In Fig. 3, for instance, if the leftmost
FIFO had 4 physical registers R, the leftmost depth-controlled register file DCF of Fig. 4
would be controlled by the control signal at the leftmost issue slot I1 so as to place its output
terminal at the second register (N-2, N=4), while the lower two registers (N, N-1) remain
unused. Thus, the control signals controlling the depth of the depth-controlled register files
DCF are part of the bit patterns in the corresponding issue slots I1 to I4.

        In summary, a programmable processor template for implementing systolic
arrays can be achieved by providing a depth-configurable memory or register file DCF at the
input terminals of each function unit FU. The depth of the depth-configurable register file
DCF is controlled e.g. by respective bits loaded in the corresponding issue slot. With this
augmentation, systolic arrays can now be mapped onto a programmable processor, e.g. a

VLIW processor, without the need for explicitly issuing operations to implement the register moves that constitute the delay lines of the array. The proposed template can be configured to implement a variety of systolic arrays. It provides for a coarse-grained reconfigurable fabric that allows for hardware-like data throughput, at the same time that it preserves the programmability of a processor.

It is to be noted that the present invention is not restricted to the preferred embodiment but can be used in any programmable or reconfigurable data processing architecture so as to implement systolic or other pipeline architectures.

CLAIMS:

1.   A processing apparatus for implementing a systolic-array-like structure, said
apparatus comprising:

a)   input means for inputting data;

b)   register means for storing said input data in a predetermined sequence;

c)   processing means for processing data received from said register means based
on control signals generated from instruction data; and

d)   register control means for controlling the depth of said register means in
accordance with said instruction data.

2.   An apparatus according to claim 1, wherein said register means comprises
distributed register files provided at input terminals of a plurality of functional units of said
processing means.

3.   An apparatus according to claim 2, wherein said distributed register files
comprise depth-configurable FIFO register files addressable for individual registers.

4.   An apparatus according to claim 3, wherein said register control means are
arranged to determine the last logical register of said FIFO register files based on control
signals derived from said instruction data.

5.   An apparatus according to any of the preceding claims, further comprising at
least one issue slot for storing said instruction data.

6.   An apparatus according to claim 5, wherein said register control means are
arranged to use a part of the bit pattern of said instruction data stored in said at least one issue
slot for controlling said depth of said register means.

7.   An apparatus according to any one of the preceding claims, wherein said programmable processing apparatus is a scalable VLIW processor or a coarse-grained reconfigurable processor.

8.   An apparatus according to any one of claims 2 to 7, wherein said distributed register files are connected to an interconnect network made up of a plurality of point-to-point connection lines.

9.   An apparatus according to claim 8, wherein said point-to-point interconnect lines have a single source.

10.   An apparatus according to claim 8, wherein said interconnect network is partially connected.

11.   A processing method for implementing a systolic-array-like structure, said method comprising:

a)   storing said input data in a register file in predetermined sequence;

b)   processing data received from said register file based on control signals generated from instruction data; and

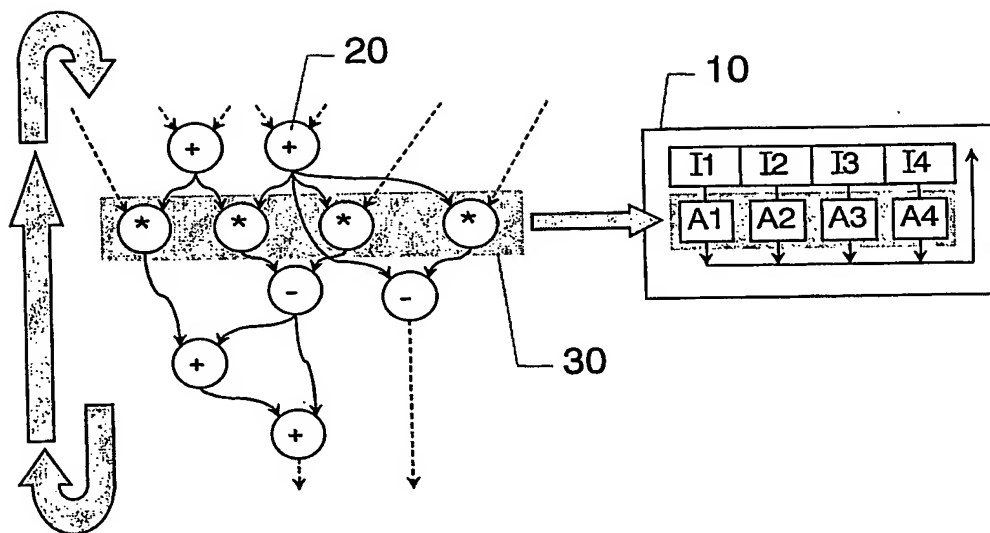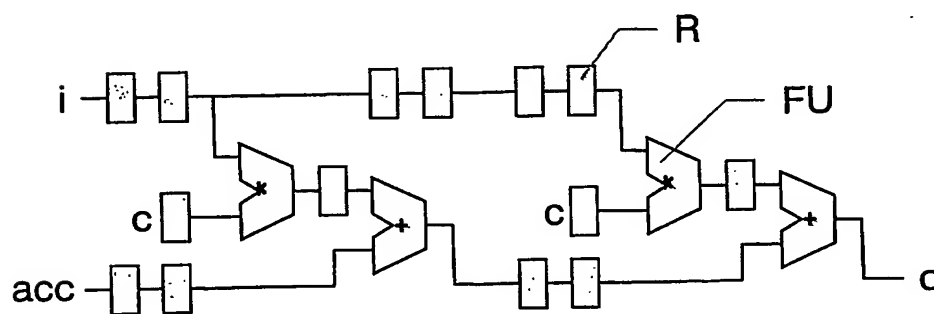c)   controlling the depth of said register file in accordance with said instruction data.
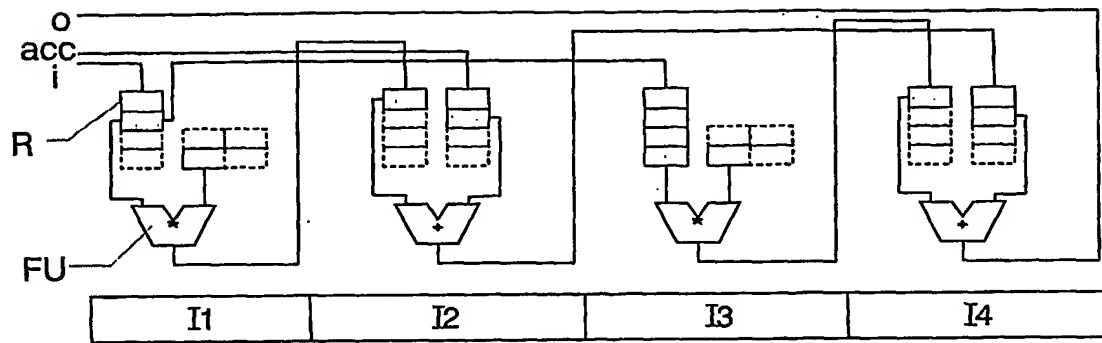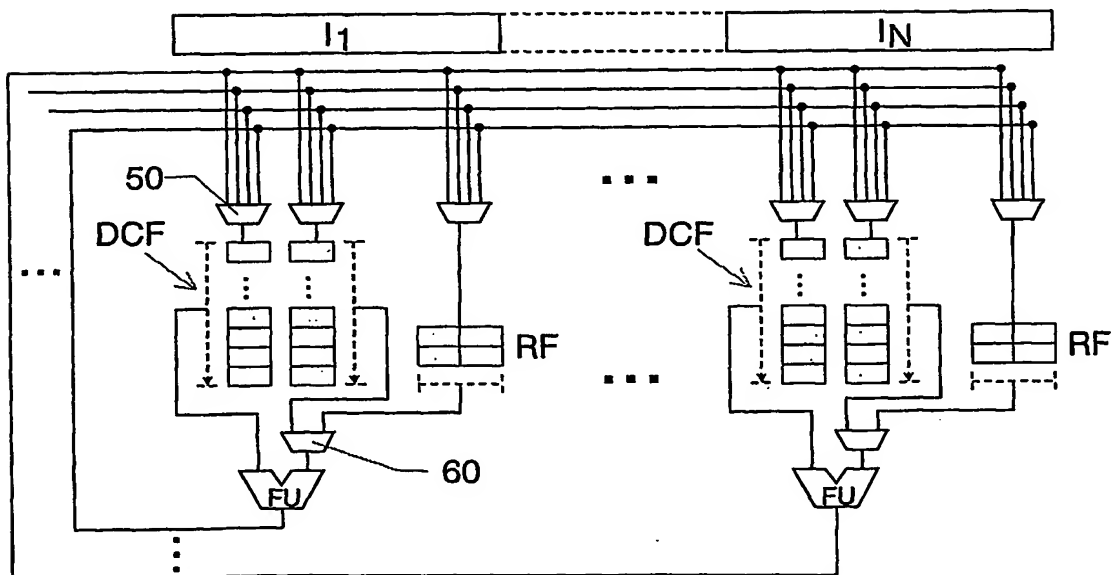
FIG.1

FIG.2

FIG.3

FIG.4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: G06F 15/80

(21) International Application Number:
PCT/IB2003/001187

(22) International Filing Date: 1 April 2003 (01.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02076649.9   25 April 2002 (25.04.2002)   EP

(71) Applicant (for all designated States except US): KONIN-KLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and
(75) Inventor/Applicant (for US only): DE OLIVEIRA KAS-TRUP PEREIRA, Bernardo [BR/NL]; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: DUIJVESTIJN, Adrianus, J.; Philips Intellectual Property & Standards, Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

[Continued on next page]

(54) Title: PROCESSING METHOD AND APPARATUS FOR IMPLEMENTING SYSTOLIC ARRAYS
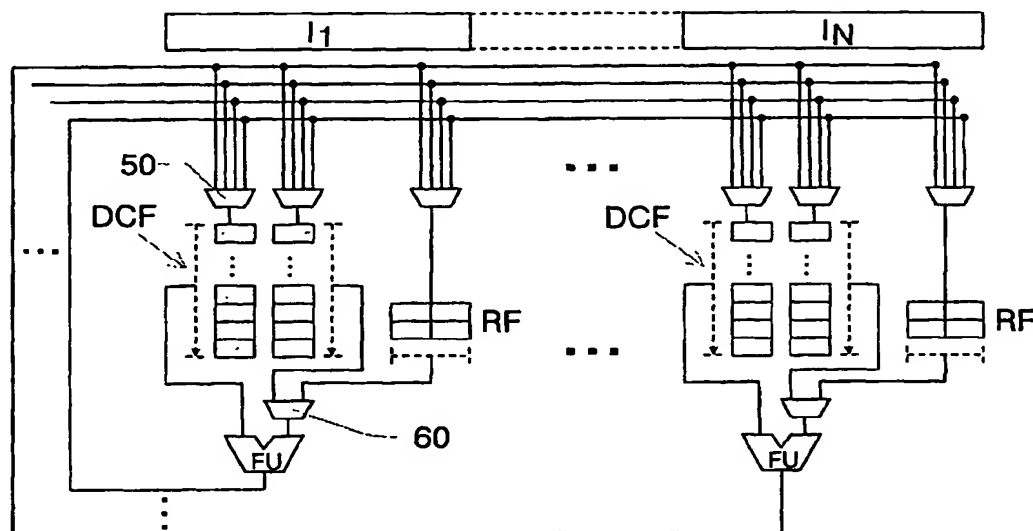
(57) Abstract: The present invention relates to a processing method and apparatus for implementing a systolic-array-like structure. Input data are stored in a depth-configurable register means (DCF) in a predetermined sequence, and are supplied to a processing means (FU) for processing said input data based on control signals generated from instruction data, 5 wherein the depth of the register means (DCF) is controlled in accordance with the instruction data. Thereby, systolic arrays can be mapped onto a programmable processor, e.g. a VLIW processor, without the need for explicitly issuing operations to implement the register moves that constitute the delay lines of the array.

WO 2003/092171 A3

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
21 October 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   G06F15/80

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, IBM-TDB, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 717 347 A (MOTOROLA INC) 19 June 1996 (1996-06-19) abstract column 6, line 37 - column 10, line 47; figure 3	1-3,5,7
A		11
A	EP 0 729 106 A (NOKIA MOBILE PHONES LTD) 28 August 1996 (1996-08-28) column 5, line 54 - column 6, line 27	4,11
	-/--	

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 August 2004	20/08/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax (+31–70) 340–3016	Michel, T

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ZAPATA E L ET AL: "A VLSI CONSTANT GEOMETRY ARCHITECTURE FOR THE FAST HARTLEY AND FOURIER TRANSFORMS" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE INC, NEW YORK, US, vol. 3, no. 1, 1992, pages 58-70, XP000262132 ISSN: 1045-9219 cited in the application the whole document —————	1,11

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0717347	A	19-06-1996	US	5673396 A	30-09-1997
			CN	1159629 A ,B	17-09-1997
			EP	0717347 A2	19-06-1996
			JP	8235850 A	13-09-1996
EP 0729106	A	28-08-1996	GB	2298109 A	21-08-1996
			DE	69625631 D1	13-02-2003
			DE	69625631 T2	11-09-2003
			EP	0729106 A1	28-08-1996
			US	5802351 A	01-09-1998